Ivip's DRTM (Distributed Real Time Mapping) system

Robin Whittle 2010-03-26

This PDF and PDFs of the illustrations can be found at http://www.firstpr.com.au/ip/ivip/drtm/.

The PDFs of the illustrations have nicer fonts than the versions in this 9 page PDF, and can be viewed horizontaly on a computer screen. There are also PowerPoint slides and .gifs of the illustrations.

Summary

DRTM is a fully distributed, non-centralised, real-time mapping system for Ivip. It could also be applied to LISP. No server needs to store the full mapping database of all MABs (Mapped Address Blocks). ISP and end-user networks with ITRs have their ITRs request mapping from a caching only Resolving Query Server (QSR) they run, - or two or three such QSRs for robustness. The QSR queries typically nearby QSA Authoritative Query Servers, which are full-database, real-time-updated, mapping servers, but only for a subset of MABs. QSAs are provided by the organisations which control MABs - ISPs do not run these.

DRTM begins with DITRs (Default ITRs in the DFZ), which involves no ITRs or any other investment by ISPs. Later, ISPs are motivated to install ITR and QSRs - which use query the QSAs at (typically) nearby DITR sites.

This is a description of the most important parts of the DRTM system, with diagrams. This is easier to read and understand than http://tools.ietf.org/html/draft-whittle-ivip-drtm .

Introduction

From its inception in mid 2007 until February 2010, Ivip's real-time mapping system was a single global system which pushed real-time mapping to full-database query servers all over the world. While each part of this was distributed, and while it was to be the result of collaboration between multiple organizations, the system was subject to two critiques:

1 - The single global system required coordination between multiple organizations.

2 - There were to be large numbers of "full-database" query servers in the networks of ISPs and larger end-user networks which ran ITRs. "Full-database" in this context means a mapping server which has a complete, real-time updated, copy of the mapping for every micronet in the Ivip system.

DRTM overcomes these concerns. It still enables end-user networks (EUNs) - or organizations they appoint - to control the mapping of micronets in real-time. This means the tunneling behavior of all ITRs handling packets addressed to a given micronet can be altered within a few seconds. (In principle the total delay time might be as low as half a second, but it best to think of one, two or three seconds). DRTM has the following advantages:

1 - No ISP or EUN (end-user network) needs to maintain query servers which receives complete feeds of real-time mapping changes, or which stores a complete copy of the mapping database.

2 - Neither do they need to run any query servers which have the full mapping database for any MABs. A Mapped Address Block (MAB) is a DFZ-advertised prefix which covers typically hundreds to tens of thousands of micronets. A micronet is an integer range of IPv4 addresses, or IPv6 /64s, which are mapped to a single ETR address. ISPs and EUNs which wish to run ITRs will run typically two or more Resolving Query Servers (QSRs), which are caching-only devices.

3 - SPI (Scalable PI) address services, including those for EUNs to scalably achieve portability, multihoming and inbound Traffic Engineering (TE) will initially be available without ISPs or EUNs making any investment, such as in ITRs or QSRs. ISPs whose EUN customers use SPI space need only ensure that their routing system accepts and forwards these EUN's outgoing packets which have SPI source addresses. In this initial phase of deployment, all packets with SPI destination addresses will be handled by DITRs - Default ITRs in the DFZ - which will be run by or on behalf of the MABOCs (MAB Operating Companies) who run each MAB. This is depicted in Figure 1 - *Stage 1: DITRS only*.

4 - As adoption of SPI space becomes more prevalent, ISPs will wish to have ITRs in their networks, for at least two reasons. Firstly so their customers will not depend on the DITRs for their outgoing packets to SPI-using EUNs. Secondly, so that when their customers send packets addressed to the SPI addresses which are used by EUNs whose ETRs are within the ISP's network, having these packets handled by internal ITRs rather than the external DITRs means that the packets do not need to leave the ISP's network or return - so saving on bandwidth to other ASes. When ISPs do adopt ITRs, they can do so gradually and need only install the ITRs and two or so QSRs, which are caching "Resolving" query servers. These query typically "nearby" QSAs (Authoritative Query Servers) at DITR sites, which are "full-database" for the subset of MABs each DITR site handles.

This document does not cover the DNS arrangements by which QSRs discover two or more typically "nearby" (a few thousand km or so) QSAs for each MAB in the Ivip system. Please refer to *http://tools.ietf.org/html/draft-whittle-ivip-drtm* for descriptions of this. In the future, there will be some kind of auto-discovery system by which ITRs can discover the QSRs they can use, and/or any intermediate QSCs (Caching Query Servers).

Below the various stages of development are described, with diagrams. IPv4 is used for convenience, but the same principles apply to IPv6.

Stage 1: DITRs only

An absolutely minimal Ivip system would consist of a single MAB, for instance 33.33.0.0 /16. This could be split up into as many as 2^{16} single IPv4 micronets, which could in principle serve this number of SPI-using EUNs (End User Networks). A more likely arrangement is that many EUNs would have one or a few micronets of one to a few IPv4 addresses, and that a smaller proposition of EUNs would have more numerous and/or large micronets. So it is reasonable to think of a single MAB such as this covering the SPI space of ~10k or more separate SPI-using EUNs.

This MAB is run by a MABOC-1, the very first MAB Operating Company. While an EUN could be its own MABOC, and use the entire MAB for its own purposes, in general, for simplicity, we assume that MABs are run by MABOCs who lease out the space within the MAB to large numbers of SPI-using EUNs.

The MAB is advertised in the DFZ and so burdens the DFZ control plane with a single prefix - yet it provides global unicast address space for ~10k or so SPI-using EUNs. So this achieves great routing scalability benefits.

The bare minimum arrangement is for MABOC-A to run a single DITR which advertises this MAB in the DFZ. All packets addressed to any of the micronets in this MAB will be forwarded to this DITR, which will encapsulate them and tunnel them to whichever ETR anywhere in the world, is specified by the mapping for the matching micronet. This single DITR approach has obvious problems with potentially high "stretch" (longer packet paths than would be required between the sending and destination hosts), with being a single-point of failure, and with a single DITR having to handle all traffic sent to these ~10k EUNs.

For MABOC-A to run this MAB in a businesslike manner, it needs DITRs in multiple locations all around the Net. This is on the reasonable assumption that its SPI-using EUN customers want to use their address space anywhere in the Net, and that the hosts which send packets to these EUN's micronets will likewise be scattered all around the Net.

So the simplest, practical, efficient instance of Ivip is a single MABOC, with a single MAB, and with DITR sites all around the Net. There's no absolute number which is required - the more the better for load sharing and reducing stretch. For instance, it would be a good start if MABOC-A established DITR sites in a dozen or so locations such as: Sydney, Singapore, Hong Kong, Tokyo, Mumbai, Moscow, Hamburg, London, Johannesburg, New York, Los Angeles and Sao Paulo.

MABOCs could be ISPs, but they need not be. Multiple MABOCs could operate independently, and each could have multiple MABs.

Each MABOC could have its own global set of DITR sites, or it could pay another organisation to run some or all the DITRs it needs to properly support its one or more MABs.

Although this discussion primarily concerns MABOCs leasing SPI space to non-mobile SPI-using EUNs, it seems likely that an early use of Ivip will be for TTR Mobility:

http://www.firstpr.com.au/ip/ivip/#mobile

A company operating a system of TTRs for its mobile end-user customers would also generally need to have TTR-sites widely dispersed around the Net, as just described for DITR sites - so that no matter where a MN (Mobile Node - AKA Mobile Host) was, there would be a TTR not too far away. A TTR company could be a MABOC, or it could be independent of any MABOCs, and each of its MN customers would provide their own one or more micronets of SPI space to use with the TTR system. The TTR company would be authorised by its customers to control the mapping of these micronets.

With multiple MABOCs of various sizes, it seems likely that most of the DITR work will be done by specialised companies we will call DSOCs - DITR Site Operating Companies. More generally, we will use the term DSOC to refer to any organization which runs a coordinated set of DITR sites around the globe. If MABOC-A ran its own DITR sites, then it would be a DSOC, just for itself. If MABOC-A also used its DITRs to support the MABs of other MABOCs, then it would be a DSOC for itself and for other MABOCs.

As a generalization of this situation, we assume the following divisions of work:

1 - An EUN may control the mapping of its micronets, but a multihomed EUN is most likely to appoint a Multihoming Monitoring and Mapping Control (MMMC) company to control the mapping of its micronets, according to the results of the MMMC company's probing of reachability of the network via its two or more ETRs.

2 - The MABOC receives these mapping change commands and ensures they are sent to the QSAs at each DITR site which supports this MAB.

3 - The DITR site may be run by the MABOC or by a separate DSOC company.

ISPs need make no investment for these early services. They do not need to install ITRs or anything else. ISPs need not run ETRs if their customers want to use SPI space, since the customer can run their own ETR on an IP address they get from the ISP (as PA space) as part of their existing Internet service.

The only thing an ISP would need to do to enable its customers to use SPI space would be to accept from these networks packets whose source address is an SPI address, and to forward these packets to their destination, including within the ISP's network or outside it via the DFZ.

Please refer to **Figure 1** which depicts an early stage of Ivip introduction. I suggest reading all the text in the diagram along with the next section.

There are multiple DSOCs, each running their own network of DITR sites. Each network of DITRs supports one or more MABs, and every MAB in the Ivip system must be supported by at least one DITR - usually by all the DITRs in one DITR network.

The diagram shows several DITR sites of DITR network A and of DITR network B. There could be dozens or in principle thousands of DITR networks. It is likely that in a full deployment there would be a dozen or more, even if there are hundreds of MABOCs and tens of thousands of MABs. This is because economies of scale would favour most MABOCs paying an existing DSOC to use support their MABs, rather than creating their own DITR network.

Figure 1 only shows, with olive lines, the raw packets sent from sending hosts H1, H5 and H7 which have SPI destination addresses. Depending on which MAB these addresses are within, and which DITRs are advertising this MABs, the packets will be forwarded out of their networks, into the DFZ, and will soon be accepted by a typically nearby DITR at some DITR site which supports the MAB.

The DITRs at these sites will either have the mapping for the matching micronet already in their cache, or will query their adjacent full-database QSA query server to obtain the mapping.

The QSAs are full-database, real-time updated, for all the MABs each DITR site handles.

Ivip does not at present specify how DSOCs and MABOCs will, or may, ensure each DITR site gets the full set of mapping changes for each MAB the site supports. This is assumed not to present any prohibitive practical problems, since only one or a few organisations must work together (such as a DSOC and multiple MABOCs) and because there is a low, finite, number of such sites. Five to a dozen would probably be fine for initial services.

The sites could use private network links for mapping changes and control, which would make them immune to any flooding or DoS attacks launched via the Internet.

All the devices mentioned here could be implemented with software on a COTS server: ITRs, DITRs, QSCs, QSRs and QSAs. ITR and DITR functions could also be added to conventional routers.

For simplicity, Figure 1 does not depict the path of the tunneled packets to the ETRs all over the world. Since there are no SPI-using networks on the left of the diagram, none of these packets will be tunneled to any network on the left of the diagram.

Please now refer to **Figure 2** and its text. This is still a Stage 1: DITRs only arrangement. The two ISPs shown in the diagram have not yet installed ITRs.

What has changed from Figure 1 is that a previously non-multihomed PA customer of ISP-F has started using SPI space (one or more micronets), and has multihomed its use of this space by getting a second Internet service from ISP-G in the same city. This customer's network is simply referred to as the EUN (End-User Network) since it is the only such network in these diagrams. Please refer to the Internet Draft for information on the EUNs relationship with the MABOC - the company which runs the MAB the EUN's micronets of SPI space are within. The EUN leases this space from the MABOC, and pays per mapping change and according to the traffic addressed to its network which flows through the DITRs. (Mapping changes are likely to cost a few cents or tens of cents at the most.)

ISP-F has not made any investment to allow its customer to use SPI space. However, in order for the EUN to be able to use this space at all, its two ISPs must accept outgoing packets whose source address is from these micronets (that is, the source address is an SPI address - it is covered by one of the MABs). The ISPs must forward these packets normally, which means to any destinations within their own networks, or otherwise out to the DFZ. Those packets emitted by the EUN, or by any other hosts or customer networks within these ISPs' networks, where the destination addresses are SPI addresses, will be forwarded to the DFZ and there to a typically nearby DITR, where they will be tunneled to an ETR. This is because in this diagram, neither of these ISPs have installed their own ITRs.

In Figure 3, they have installed ITRs, and these ITRs handle packets whose destination address matches any of the MABs.

In Figure 2 this has not occurred. While ISP-F is happy to retain its EUN customer with the new arrangements - its use of SPI space - there is a problem for ISP-F, especially if many of its customers do the same thing. Assuming packets are sent from hosts in the ISP's network - including hosts of its customers, (including those customers like the EUN, which are using SPI space themselves) - and these packets are addressed to SPI addresses used by any of the ISP's customers, then these packets will leave the ISP's network, find their way to a DITR, and return to the ISP's network. This involves extra cost for the ISP, since it uses the expensive upstream links.

This can be seen with the olive line from H1 to DITR-A1, and with the orange broken line returning the packet, in a tunneled form, to the ISP's network, and then to the ETR inside the EUN network. The ETR detunnels it and forwards it internally (short olive line) to the destination host.

To avoid this, both ISPs install ITRs, and two or more QSRs to resolve the mapping queries of these ITRs. Two or more QSRs are used for load sharing and robustness, but for simplicity, only one is shown in these diagrams.

Phase 2: Add ITRs with caching QSRs

Please refer to **Figure 3** and its text. Now the packet sent from H1 to H5 does not need to leave ISP-F's network - since ISP-F has installed ITRs and the QSRs it needs to support them.

Not shown are optional caching QSCs which can be between the ITRs an the QSRs. These reduce the query load of QSRs, and reduce the number of Map Update messages the QSRs need to send, to the extent that multiple ITRs served by the one QSC need the same mapping, which will typically be the case for a significant fraction of the map requests.

As an increasing number of ISPs do this, there is less load on the DITRs. So MABOCs are motivated to ensure their DITR sites have reliable, fast-responding, QSAs - to encourage ISPs to run their own ITRs and QSRs.

It is not absolutely essential that all QSAs be at DITR sites, but for this discussion I am assuming this.

The EUN has also installed ITRs. Ivip allows for sending hosts (on SPI or conventional addresses, but not behind NAT) to have an ITR function built into their networking software. So there may be essentially no cost in installing ITRs in the EUN. These ITRs do need to get mapping from one or ideally more QSRs. The EUN could install its own QSRs, but in this case, its ITRs query directly the QSRs of its two ISPs. A more likely scenario would be the EUN installing one or more caching QSC query servers which in turn directed their queries to the QSRs of the two ISPs - directly or by further levels of QSC in those ISP networks.

Multiple levels of QSC will not significantly increase the time it takes an ITR to obtain mapping - which will usually be in the 10 to 60 msec range. This is assumed not to be a significant delay for any human users or for any application protocols.

The DITR are still operating and handling packets sent from networks without ITRs, but over time, it is reasonable to expect many or most ISPs and larger EUNs to install their own ITRs. Having internal ITRs ensures packets are tunneled correctly, without having to rely on potentially congested DITRs. On the other hand, MABOCs will be highly motivated to ensure their DITRs are not congested, because if they are, then their SPI leasing EUN customers will lose some packets which are addressed to their micronets.

Conclusion

I hope this discussion and these diagrams make it easy to understand how DRTM is a staged process, in which ISPs make no initial investment and the initiative and investment is undertaken by MABOCs and their EUN customers.

The system provides real-time control of mapping to all ITRs which need it, and as described in Ivip-arch, this leads to many advantages and simplifications compared to Core-Edge Separation architectures such as LISP and IRON-RANGER which do not attempt to give EUNs (or whoever the EUN appoints) direct, real-time control of the tunneling behavior of ITRs.

Ivip DRTM Stage 1: DITRs only

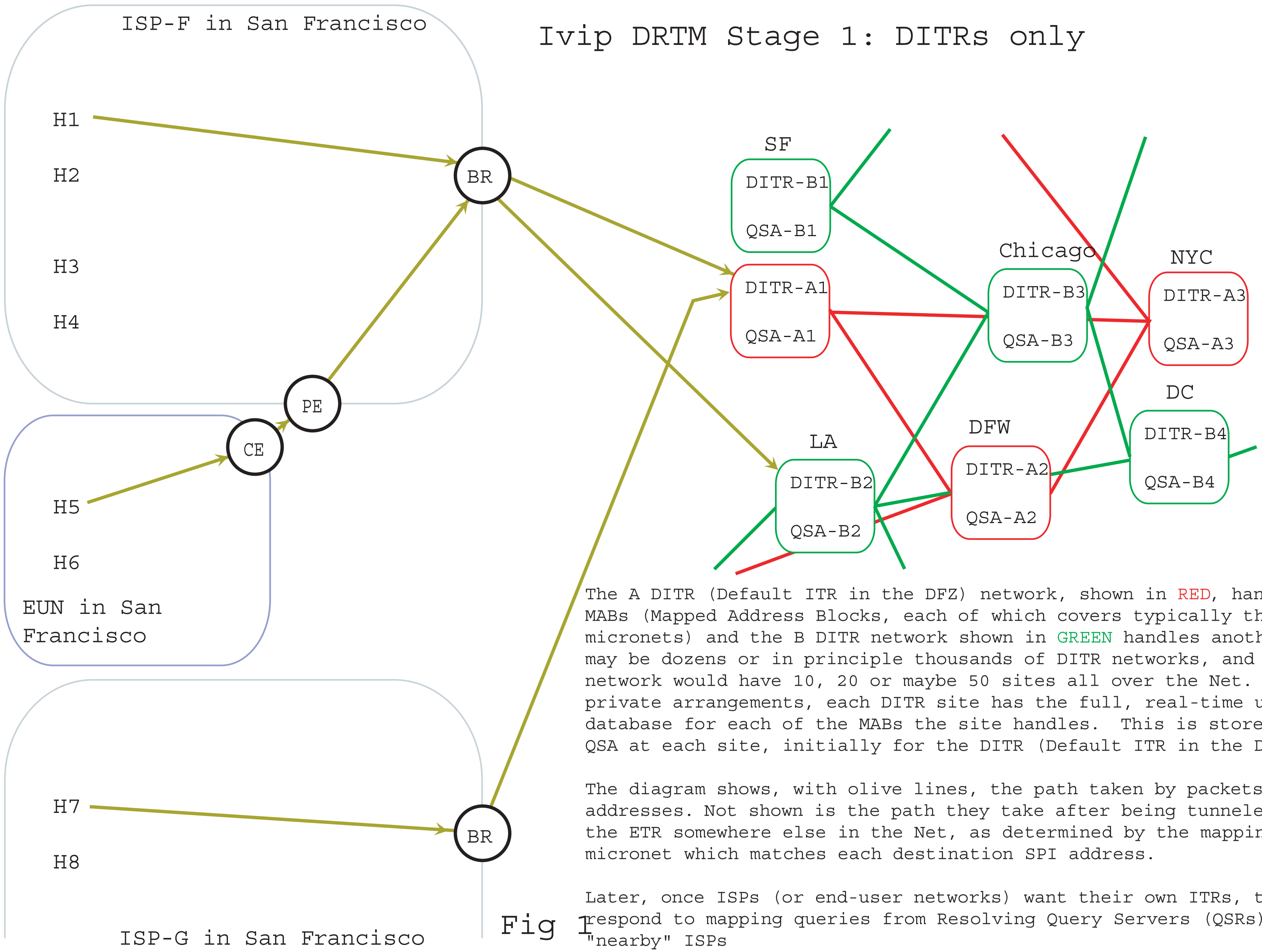


Fig 1

The A DITR (Default ITR in the DFZ) network, shown in RED, handles one set of MABs (Mapped Address Blocks, each of which covers typically thousands of micronets) and the B DITR network shown in GREEN handles another set. There may be dozens or in principle thousands of DITR networks, and ideally each network would have 10, 20 or maybe 50 sites all over the Net. By their own private arrangements, each DITR site has the full, real-time updated, mapping database for each of the MABs the site handles. This is stored in the Authoritative QSA at each site, initially for the DITR (Default ITR in the DFZ) there to query.

The diagram shows, with olive lines, the path taken by packets addressed to SPI addresses. Not shown is the path they take after being tunneled by the DITRs to the ETR somewhere else in the Net, as determined by the mapping for the micronet which matches each destination SPI address.

Later, once ISPs (or end-user networks) want their own ITRs, the QSAs will also respond to mapping queries from Resolving Query Servers (QSRs) in typically "nearby" ISPs

Ivip DRTM Stage 1: DITRs only - but a customer network uses SPI space and this motivates its ISPs to install ITRs.

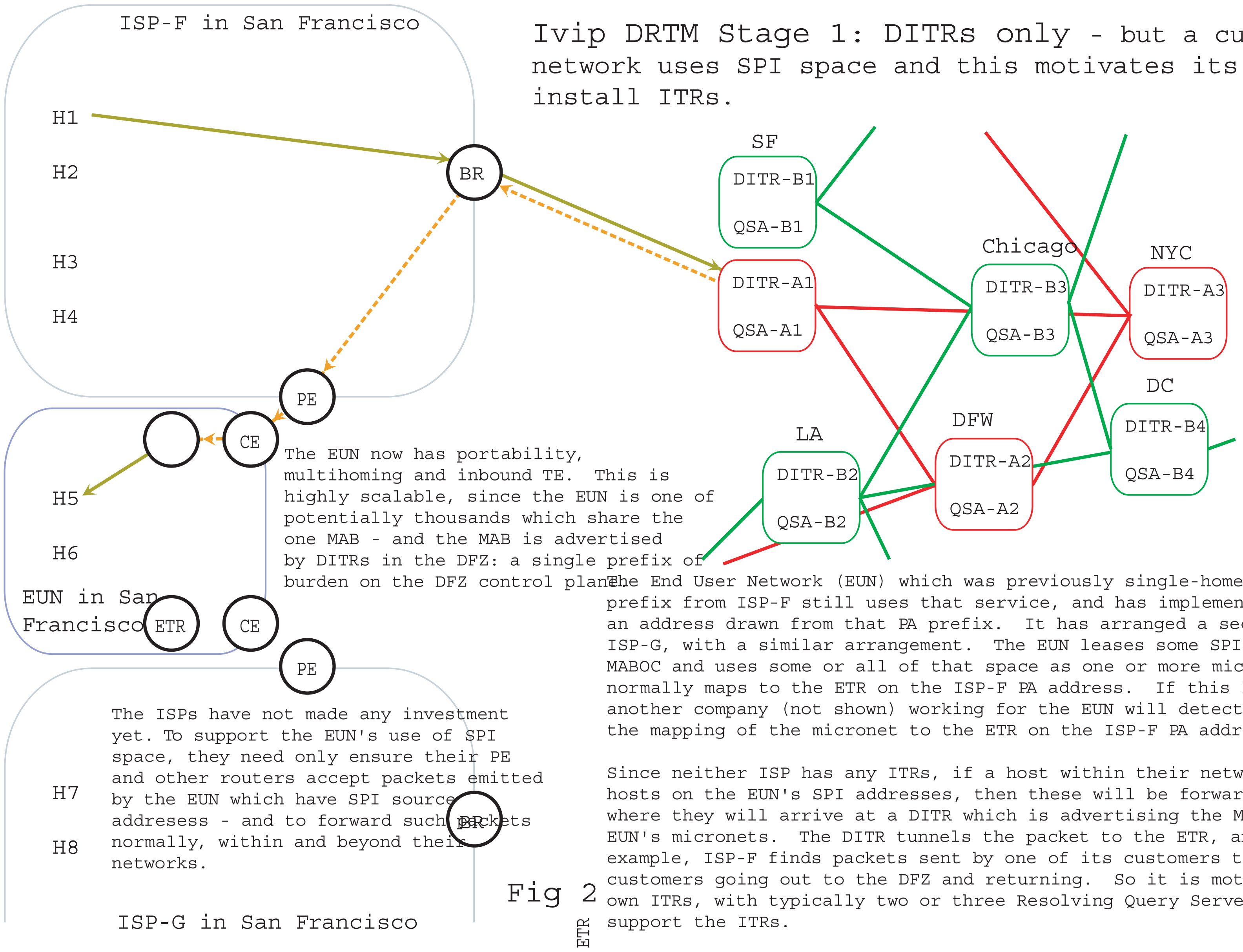


Fig 2

The EUN now has portability, multihoming and inbound TE. This is highly scalable, since the EUN is one of potentially thousands which share the one MAB - and the MAB is advertised by DITRs in the DFZ: a single prefix of burden on the DFZ control plane.

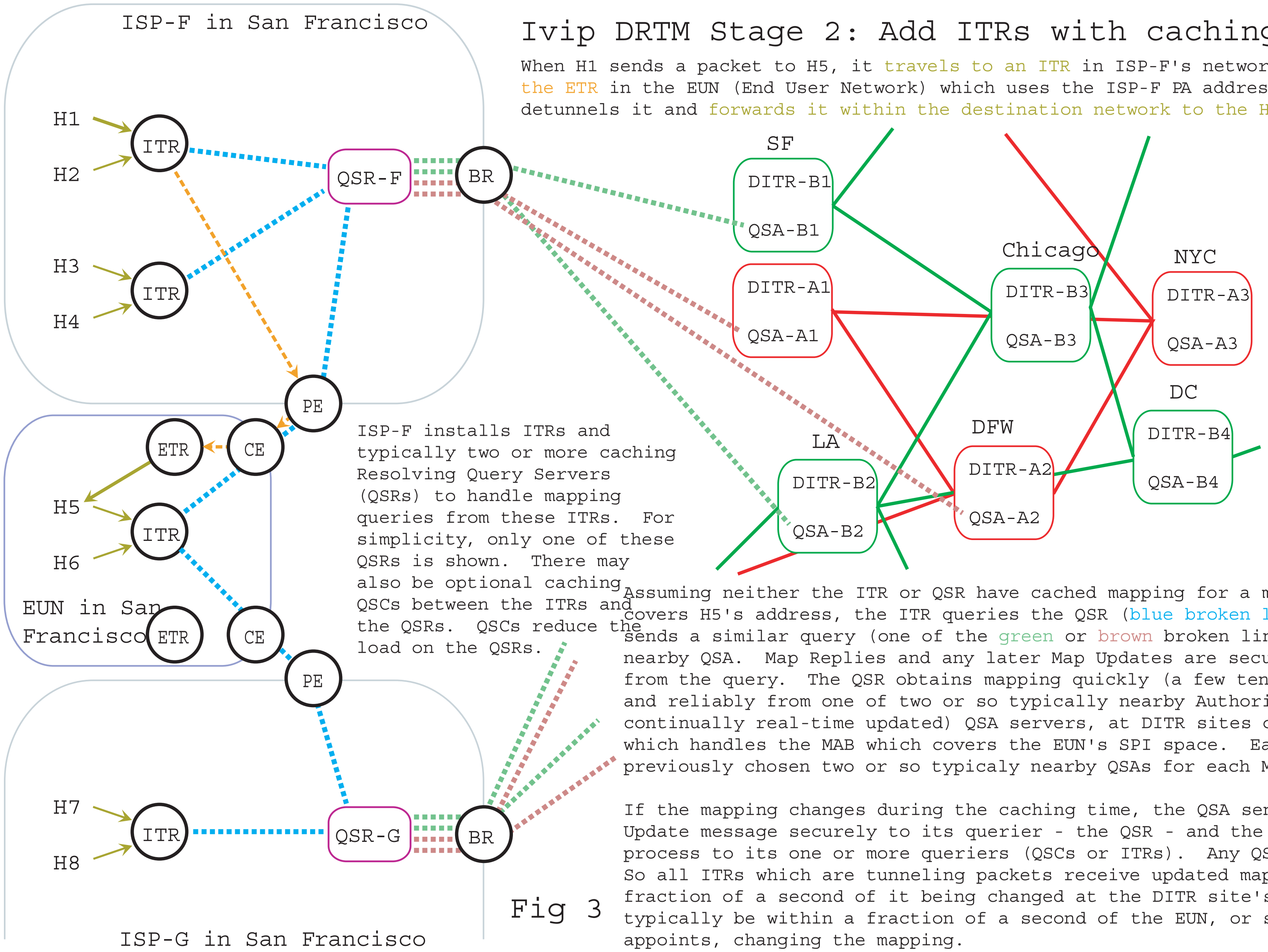
The ISPs have not made any investment yet. To support the EUN's use of SPI space, they need only ensure their PE and other routers accept packets emitted by the EUN which have SPI source addresess - and to forward such packets normally, within and beyond their networks.

The End User Network (EUN) which was previously single-homed on a PA prefix from ISP-F still uses that service, and has implemented its own ETR on an address drawn from that PA prefix. It has arranged a second service from ISP-G, with a similar arrangement. The EUN leases some SPI space from a MABOC and uses some or all of that space as one or more micronets which it normally maps to the ETR on the ISP-F PA address. If this link to ISP-F fails, another company (not shown) working for the EUN will detect this and change the mapping of the micronet to the ETR on the ISP-F PA address.

Since neither ISP has any ITRs, if a host within their network sends packets to hosts on the EUN's SPI addresses, then these will be forwarded to the DFZ where they will arrive at a DITR which is advertising the MAB which covers the EUN's micronets. The DITR tunnels the packet to the ETR, and in this example, ISP-F finds packets sent by one of its customers to another of its customers going out to the DFZ and returning. So it is motivated to install its own ITRs, with typically two or three Resolving Query Servers, QSRs to support the ITRs.

Ivip DRTM Stage 2: Add ITRs with caching QSRs

When H1 sends a packet to H5, it travels to an ITR in ISP-F's network, which tunnels it to the ETR in the EUN (End User Network) which uses the ISP-F PA address. The ETR detunnels it and forwards it within the destination network to the H5 destination host.



ISP-F installs ITRs and typically two or more caching Resolving Query Servers (QSRs) to handle mapping queries from these ITRs. For simplicity, only one of these QSRs is shown. There may also be optional caching QSCs between the ITRs and the QSRs. QSCs reduce the load on the QSRs.

Fig 3

Assuming neither the ITR or QSR have cached mapping for a micronet which covers H5's address, the ITR queries the QSR (blue broken line), and the QSR sends a similar query (one of the green or brown broken lines) to a typically nearby QSA. Map Replies and any later Map Updates are secured by a nonce from the query. The QSR obtains mapping quickly (a few tens of milliseconds) and reliably from one of two or so typically nearby Authoritative (full-database, continually real-time updated) QSA servers, at DITR sites of the DITR network which handles the MAB which covers the EUN's SPI space. Each QSR has previously chosen two or so typicaly nearby QSAs for each MAB.

If the mapping changes during the caching time, the QSA sends a Map Update message securely to its querier - the QSR - and the QSR repeats this process to its one or more queriers (QSCs or ITRs). Any QSCs do the same. So all ITRs which are tunneling packets receive updated mapping within a fraction of a second of it being changed at the DITR site's QSA - which will typically be within a fraction of a second of the EUN, or some company it appoints, changing the mapping.